CLAIM AMENDMENTS

Please amend the claims by canceling claims 12-16 and 44-48 without prejudice, as indicated on the following listing of all the claims in the present application after this Amendment:

1. (Original) A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

receiving a packet;

parsing application content from said packet;

maintaining a set of server load metrics associated with the servers among said group of servers;

selecting a destination server for said packet from among said group of servers, said selection being dependent on said set of server load metrics;

assigning to said packet a priority dependent on said parsed application content; and

transmitting said packet to the destination server according to a transmitting schedule dependent on said priority.

2. (Original) A method of switching packets to a server among a group of servers as in claim 1, further comprising:

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content, and wherein

said destination server selection is restricted to said group of eligible servers.

3. (Original) A method of switching packets to a server among a group of servers, as in claim 1, further comprising:

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

said selecting a destination server is also dependent on said estimated application load.

4. (Original) A method of switching packets to a server among a group of servers as in of claim 1, wherein said transmitting schedule is also dependent on said set of server load metrics.

5. (Original) A method of switching packets to a server among a group of servers as in claim1, wherein said application content includes that from a header for Hypertext Transfer Protocol.

6. (Original) A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

receiving a packet;

parsing application content from said packet;

maintaining a set of server load metrics associated with the servers among said group of servers;

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content;

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content;

selecting a destination server for said packet from among said group of eligible servers, said selection being dependent on said set of server load metrics and said estimated application load;

assigning to said packet a priority dependent on said parsed application content; and

transmitting said packet to the destination server according to a transmitting schedule dependent on said priority and said set of server load metrics.

7. (Original) A method of switching packets to a server among a group of servers as in claim 6, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

8. (Original) A method of switching packets to a server among a group of servers as in anyone of claims 1-7, wherein:

the priority assigned to said packet is one of a first priority or a second priority lower than said first priority; and

said transmitting schedule is such that whenever the destination server has a workload above a first predetermined level, a first priority packet is transmitted thereto without delay while a second priority packet is held back.

9. (Original) A method of switching packets to a server among a group of servers as in claim 8, wherein:

said transmitting schedule is such that whenever the destination server has a workload below said first predetermined level, a first priority packet or a second priority packet is transmitted thereto without delay.

10. (Original) A method of switching packets to a server among a group of servers as in anyone of claims 1-7, wherein:

the priority assigned to said packet is one of a hierarchy of priorities ranging from a highest priority to a lowest priority; and said method further comprising:

providing a set of predetermined workload levels for the destination server, said set of predetermined workload levels ranging from a highest level corresponding to the highest priority to a lowest level corresponding to the lowest priority; and wherein

said transmitting schedule is such that whenever the destination server has a workload between a first of said levels and a second of said levels less than said first of said levels, a packet assigned the priority corresponding to said first of said levels is transmitted to the destination server without delay while a packet assigned the priority corresponding to said second of said levels is held back.

11. (Original) A method of switching packets to a server among a group of servers as in claim 10, wherein:

the priority assigned to said packet includes one of zero priority; and

said transmitting schedule is such that a packet with zero priority is never transmitted to the destination server.

12. **(Canceled)** A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

receiving a packet;

parsing application content from said packet;

assigning to said packet a priority dependent on said parsed application content; and

dropping the packet if its assigned priority corresponds to a predetermined type.

13. **(Canceled)** A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

receiving a packet;

parsing application content from said packet; and

enforcing security by reference to an access control list dependent on said parsed application content.

14. **(Canceled)** A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, as in claim 13, wherein said enforcing of security includes dropping the packet if its parsed application content is of a first predetermined type.

15. **(Canceled)** A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, as in claim 13, wherein said enforcing of security includes redirecting the packet to a predetermined location if its parsed application content is of a second predetermined type.

16. **(Canceled)** A method of switching packets to a server among a group of servers as in claim 12, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

17. (Original) A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

a buffer memory for storing individual packets;

a content class table defining a set of content classes with respect to a corresponding set of application content;

a class policy table associating a priority to a given content class;

a server table providing a set of server load metrics associated with the servers among said group of servers;

a packet classifier for assigning a content class to a packet, said packet classifier parses the application content from said packet and determines said content class by reference to said content class table for said parsed application content;

a load balancer for selecting a destination server from among said group of servers based on a set of server load metrics by reference to said server table for said group of servers;

a Quality-of-Service controller for scheduling transmission of said packet to the destination server according to a priority assigned to said packet by reference to said class policy table for said content class; and

a transmitter, responsive to said load balancer and said Quality-of-Service controller for releasing said packet from said buffer to the destination server in the order the packet is scheduled.

18. (Original) A switch for switching packets to a server among a group of servers as in claim 17, further comprising:
a server group table associating a content class to a group of eligible servers and wherein

said load balancer selects a destination server from among said group of eligible servers.

19. (Original) A switch for switching packets to a server among a group of servers as in claim 17, further comprising:

an application load table associating a content class to an estimated application load on the selected server; and wherein:

the selected server has a server load metric that is dependent on said estimated application load.

20. (Original) A switch for switching packets to a server among a group of servers as in claim 17, wherein:

said transmission schedule of said packet to the destination server is also based on said set of server load metrics.

21. (Original) A switch for switching packets to a server among a group of servers as in claim 17, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

22. (Original) A switch for switching a packet to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

a buffer memory for storing said packet;

a content class table defining a set of content classes with respect to a corresponding set of application content;

a class policy table associating a priority to a given content class;

a server group table associating a content class to a group of eligible servers;

an application load table associating a content class to an estimated application load on the selected server;

a server table providing a set of server load metrics associated with the servers among said group of servers;

a packet classifier for assigning a content class to said packet, said packet classifier parses the application content from said packet and determines said content class by reference to said content class table for said parsed application content;

a load balancer for selecting a destination server from among said group of servers, said selecting being restricted to a group of eligible servers by reference to said server group table for said content class, and said selecting being based on said set of server load metrics by reference to said server table for said group of eligible servers and based on an estimated application load by reference to said application load table for said content class;

a Quality-of-Service controller for scheduling transmission of said packet to the destination server according to a priority assigned to said packet by reference to said class policy table for said content class and said set of server load metrics; and

a transmitter, responsive to said load balancer and said Quality-of-Service controller for releasing said packet from said buffer to the destination server in the order the packet is scheduled.

23. (Original) A switch for switching packets to a server among a group of servers as in claim 22, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

24. (Original) A switch for switching packets to a server among a group of servers as in claims 17-22, wherein

the priority assigned to said packet is one of a first priority or a second priority lower than said first priority; and

said transmitting schedule is such that whenever the destination server has a workload above a first predetermined level, a first priority packet is transmitted thereto without delay while a second priority packet is held back.

25. (Original) A switch for switching packets to a server among a group of servers as in claim 24, wherein:

said transmitting schedule is such that whenever the destination server has a workload below said first predetermined level, a first priority packet or a second priority packet is transmitted thereto without delay.

26. (Original) A switch for switching packets to a server among a group of servers as in anyone of claims 17-22, wherein:

the priority assigned to said packet is one of a hierarchy of priorities ranging from a highest priority to a lowest priority; and said switch further comprising:

a set of predetermined workload levels for the destination server, said set of predetermined workload levels ranging from a highest level corresponding to the highest priority to a lowest level corresponding to the lowest priority; and wherein

said transmitting schedule is such that whenever the destination server has a workload between a first of said levels and a second of said levels less than said first of said levels, a packet assigned the priority corresponding to said first of said levels is transmitted to the destination server without delay while a packet assigned the priority corresponding to said second of said levels is held back.

27. (Original) A switch for switching packets to a server among a group of servers as in claim 26, wherein:

the priority assigned to said packet includes one of zero priority; and

said transmitting schedule is such that a packet with zero priority is never transmitted to the destination server.

28. (Original) A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

a buffer memory for storing individual packets;

a content class table defining a set of content classes with respect to a corresponding set of application content;

a class policy table associating a priority to a given content class;

a packet classifier for assigning a content class to a packet, said packet classifier parses the application content from said packet and determines said content class by reference to said content class table for said parsed application content; and

a transmit controller for dropping the packet if its assigned priority corresponds to a predetermined type.

29. (Original) A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

a buffer memory for storing individual packets;

a content class table defining a set of content classes with respect to a corresponding set of application content;

an access control list table associating security access instructions to a given content class;

a packet classifier for assigning a content class to a packet, said packet classifier parses the application content from said packet and determines said content class by reference to said content class table for said parsed application content; and

a transmit controller responsive to said access control list table for enforcing securing access of the packet.

30. (Original) A switch for switching packets to a server among a group of servers as in claim 29, wherein said transmit controller drops the packet when its parsed application content is of a first content class.

31. (Original) A switch for switching packets to a server among a group of servers as in claim 29, wherein said transmit controller redirects the packet to a predetermined location when its parsed application content is of a second content class.

32. (Original) A switch for switching packets to a server among a group of servers as in claim 28, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

33. (Original) A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

means for receiving a packet;

means for parsing application content from said packet;

means for maintaining a set of server load metrics associated with the servers among said group of servers;

means for selecting a destination server for said packet from among said group of servers, said selection being dependent on said set of server load metrics;

means for assigning to said packet a priority dependent on said parsed application content; and

means for transmitting said packet to the destination server according to a transmitting schedule dependent on said priority.

34. (Original) A switch for switching packets to a server among a group of servers as in claim 33, further comprising:

means for determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content, and wherein

said destination server selection is restricted to said group of eligible servers.

35. (Original) A switch for switching packets to a server among a group of servers, as in claim 33, further comprising:

means for determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

said selecting a destination server is also dependent on said estimated application load.

36. (Original) A switch for switching packets to a server among a group of servers as in of claim 33, wherein said transmitting schedule is also dependent on said set of server load metrics.

37. (Original) A switch for switching packets to a server among a group of servers as in anyone of claim 33, wherein said application content includes that from a header for Hypertext Transfer Protocol.

38. (Original) A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

means for receiving a packet;

means for parsing application content from said packet;

means for maintaining a set of server load metrics associated with the servers among said group of servers;

means for determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content;

means for determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content;

means for selecting a destination server for said packet from among said group of eligible servers, said selection being dependent on said set of server load metrics and said estimated application load;

means for assigning to said packet a priority dependent on said parsed application content; and

means for transmitting said packet to the destination server according to a transmitting schedule dependent on said priority and said set of server load metrics.

39. (Original) A switch for switching packets to a server among a group of servers as in claim 38, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

40. (Original) A switch for switching packets to a server among a group of servers as in anyone of claims 33-39, wherein:

the priority assigned to said packet is one of a first priority or a second priority lower than said first priority; and

said transmitting schedule is such that whenever the destination server has a workload above a first predetermined level, a first priority packet is transmitted thereto without delay while a second priority packet is held back.

41. (Original) A switch for switching packets to a server among a group of servers as in claim 40, wherein:

said transmitting schedule is such that whenever the destination server has a workload below said first predetermined level, a first priority packet or a second priority packet is transmitted thereto without delay.

42. (Original) A switch for switching packets to a server among a group of servers as in anyone of claims 33-39, wherein:

the priority assigned to said packet is one of a hierarchy of priorities ranging from a highest priority to a lowest priority; and said method further comprising:

means for providing a set of predetermined workload levels for the destination server, said set of predetermined workload levels ranging from a highest level corresponding to the highest priority to a lowest level corresponding to the lowest priority; and wherein

said transmitting schedule is such that whenever the destination server has a workload between a first of said levels and a second of said levels less than said first of said levels, a packet assigned the priority corresponding to said first of said levels is transmitted to the destination server without delay while a packet assigned the priority corresponding to said second of said levels is held back.

43. (Original) A switch for switching packets to a server among a group of servers as in claim 42, wherein:

the priority assigned to said packet includes one of zero priority; and

said transmitting schedule is such that a packet with zero priority is never transmitted to the destination server.

44. **(Canceled)**   A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

means for receiving a packet;

means for parsing application content from said packet;

means for assigning to said packet a priority dependent on said parsed application content; and

means for dropping the packet if its assigned priority corresponds to a predetermined type.

45. **(Canceled)**   A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, comprising:

means for receiving a packet;

means for parsing application content from said packet; and

means for providing application layer security by enforcing access control dependent on said parsed application content.

46. **(Canceled)** A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, as in claim 45, wherein:

said means for providing application layer security includes dropping the packet if its parsed application content is of a first predetermined type.

47. **(Canceled)** A switch for switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, as in claim 45, wherein:

said means for providing application layer security includes redirecting the packet to a predetermined location if its parsed application content is of a second predetermined type.

48. **(Canceled)** A switch for switching packets to a server among a group of servers as in claim 44, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

49. (Original) A method of switching packets to a server among a group of servers including a newly added server participating in load balancing, comprising:

maintaining a set of server load metrics associated with the servers among said group of servers, wherein a load metric associated with a server is dependent on the number of connections currently made hereto, and a disparity initially exists in server load metrics between said newly added server and that of other servers among said group of servers;

multiplying a time-varying factor to the server load metric for said newly added server such that said disparity is substantially reduced to below a predetermined level over a predefined period of time;

selecting for a packet a destination server having a minimum server load metric among said group of servers; and

transmitting said packet to the destination server.

50. (Original) A method of switching packets as in claim 49, wherein:

said time-varying factor decreases from a first value substantially above unity to unity over said predefined period of time.

51. (Original) A method of switching packets as in claim 50, wherein:

said first value is given by 2^k, where k is an integer, and is reduced by a factor of two at a time until said first value becomes unity.

52. (Original) A method of switching packets to a server among a group of servers including a newly added server participating in load balancing, comprising:

maintaining a set of server load metrics associated with the servers among said group of servers, wherein a load metric associated with a server is dependent on the number of connections currently made hereto, and a disparity initially exists in server load metrics between said newly added server and that of other servers among said group of servers;

multiplying a factor of 2^k, with k being an integer, to the server load metric for said newly added server;

reducing said factor of 2^k by a factor of two at predefined periods until unity is reached;

selecting for said packet a destination server having a minimum server load metric among said group of servers; and

transmitting said packet to the destination server.

53. (Original) A method of switching packets as in anyone of claims 49-52, wherein:

a server weight is associated with each server to indicate a capacity capability of each server; and

the load metric of each server is weighted by its associated server weight.

54. (Original) A method of switching packets as in anyone of claims 49-52, wherein individual packets carry content associated with an application, said method further comprising:

receiving a packet;

parsing application content from said packet;

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content, and wherein

said destination server selection is restricted to said group of eligible servers.

55. (Original) A method of switching packets as in anyone of claims 49-52, further comprising:

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

the server load metric for the destination server is also dependent on said estimated application load.

56. (Original) A method of switching packets as in claim 55, wherein said application content includes that from a header for Hypertext Transfer Protocol.

57. (Original) A method of switching packets as in anyone of claims 49-52, said method comprising:

receiving a packet;

parsing application content from said packet;

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content;

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

the server load metric for the destination server is also dependent on said estimated application load.

58. (Original) A method of switching as in claim 57, wherein said application content includes that from a packet header for Hypertext Transfer Protocol.

59. (Original) A method of switching packets to a server among a group of servers, wherein individual packets carry content associated with an application, said method comprising:

maintaining a set of server load metrics associated with the servers among said group of servers,

receiving a packet during a TCP session;

parsing application content from said packet;

updating said set of server load metrics only when the TCP session is not in an initial handshaking phase, said updating incorporating an estimated application load determined with reference to the application content parsed from said packet;

selecting a destination server for said packet from among said group of servers, said selection being dependent on said set of server load metrics;

assigning to said packet a priority dependent on application content parsed from said packet only when the TCP session is not in an initial handshaking phase, otherwise assigning to said packet a default priority; and

transmitting said packet to the destination server according to a transmitting schedule dependent on said priority

60. (Original) A method of switching packets to a server among a group of servers as in claim 59, wherein said default priority is a highest among a group of possible priority types.

61. (Original) A method of switching packets to a server among a group of servers as in claim 59, further comprising:

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content, and wherein

said destination server selection is restricted to said group of eligible servers.

62. (Original) A method of switching packets to a server among a group of servers, as in claim 59, further comprising:

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

said selecting a destination server is also dependent on said estimated application load.

63. (Original) A method of switching packets to a server among a group of servers as in of claim 59, wherein said transmitting schedule is also dependent on said set of server load metrics.

64. (Original) A method of switching packets to a server among a group of servers as in claims 59, wherein said application content includes that from a header for Hypertext Transfer Protocol.

65. (Original) A method of switching packets to a server among a group of servers as in claims 60, wherein said application content includes that from a header for Hypertext Transfer Protocol.

66. (Original) A method of switching packets to a server among a group of servers as in claims 61, wherein said application content includes that from a header for Hypertext Transfer Protocol.

67. (Original) A method of switching packets to a server among a group of servers as in claims 59, wherein said application content includes that from a header for Hypertext Transfer Protocol.

68. (Original) A method of switching packets to a server among a group of servers as in claims 63, wherein said application content includes that from a header for Hypertext Transfer Protocol.

69. (Original) A method of switching packets to a server among a group of servers as in claim 60, further comprising:

determining a group of eligible servers among said group of servers, said determination being dependent on said parsed application content, and wherein

said destination server selection is restricted to said group of eligible servers.

70. (Original) A method of switching packets to a server among a group of servers, as in claim 60, further comprising:

determining an estimated application load on the destination server, said estimated application load being dependent on said parsed application content; and wherein

said selecting a destination server is also dependent on said estimated application load.

71. (Original) A method of switching packets to a server among a group of servers as in of claim 60, wherein said transmitting schedule is also dependent on said set of server load metrics.

72. (Original) A method of switching packets to a server among a group of servers as in claim 69, wherein said application content includes that from a header for Hypertext Transfer Protocol.

73. (Original) A method of switching packets to a server among a group of servers as in claim 70, wherein said application content includes that from a header for Hypertext Transfer Protocol.

74. (Original) A method of switching packets to a server among a group of servers as in claim 71, wherein said application content includes that from a header for Hypertext Transfer Protocol.